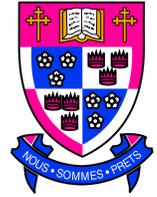




# Optimization Methods for Binary Sequences – The Merit Factor Problem

Ron Ferguson, Joshua Knauer SFU  
 MOCAA Project Leader: Peter Borwein  
 MITACS 6th Annual Conference, May, 2005



## Abstract

Optimization procedures for binary sequences, applied in particular to the merit factor problem, have been a subject of much interest in combinatorial optimization, communications engineering, and analytic number theory. Exact methods have confirmed optimal sequences up to length 60, but this is not far enough to adequately make projections. We combine both stochastic and non-stochastic search methods to obtain what we consider comprehensive results to length 90 in general and to length 180 for restriction to skew-symmetric sequences and develop statistical arguments to support this claim.

## 1 Merit Factor

The merit factor is a term introduced by communications engineers as a quality measure for binary sequences in the form  $\{a_1, a_2, \dots, a_n\}$  where  $a_i = \pm 1$ . The  $k$ -shift sidelobe, or autocorrelation coefficient, is defined by

$$c_k = \sum_{i=1}^{n-k} a_i a_{i+k}.$$

Its energy is given by  $c_k^2$  and measures self-interference of a signal. The merit factor is the ratio of the energy in the main (0-shift) sidelobe to the sum of the energies in the remaining sidelobes:

$$F = \frac{n^2}{2 \left( \sum_{k=1}^{n-1} c_k^2 \right)}.$$

The minimum possible sidelobe energies are either 0 or 1 depending on parity. *Barker* sequences are optimal in this way, and were investigated for use as radar signals. However, the longest known has length 13, and it is generally believed that there are no more at longer lengths. A stronger conjecture, namely, that the merit factor is absolutely bounded, is widely believed. Many suspect that the merit factor of the length 13 Barker sequence (14.08) is the largest possible.

A *skew-symmetric* sequence is partially optimal, having a 0 value for all even length sidelobe sums. It is formed by the interleaving of an odd length symmetric sequence with an even length anti-symmetric sequence.

## 2 Search Methods

An exhaustive search would examine all sequences and have complexity  $2^N$  for regular sequences and  $2^{(N+1)/2}$  for skew symmetric.

### 2.1 Sequence pairing

This is a method used by Golay and Harris to find skew-symmetric sequences of high merit. The symmetric and antisymmetric sequences with merit factor greater than a set bound were first collected. Through exhaustively interleaving appropriate pairs from these collections, sequences with high merit factor were obtained up to

length 121. In many cases the optimal skew-symmetric sequence was obtained, although the method doesn't confirm this.

### 2.2 Branch and bound

This method has confirmed optimality of sequences for lengths up to 60, which is about the limit of current methods and equipment. As developed by Mertens and others it uses the idea that as a sequence is formed starting from the ends and progressing inward, the values of successive sidelobes are established. Branching occurs when, without completing the sequence, the known energies in the shorter sidelobes already exceed an established bound. Statistical analysis suggest the complexity reduces to  $1.85^N$ .

### 2.3 Stochastic optimization

For an entering sequence an entry is chosen at random and its sign reversed. A decision is then made to either accept or undo the change according to the change in merit factor. This continues until no further changes of this kind are acceptable. Straight hill climbing, where the criterion is an increase in the merit factor are not the most efficient. Methods employing variations of synthetic annealing or the great deluge algorithm perform better. This has been used to good sequences beyond lengths beyond 60, but doesn't confirm optimality.

### 2.4 Comprehensive searches of clusters

The basic premise here is that a sequence of high merit will be contained in a cluster of such sequences within a short Hamming distance and not restricted to the same length. Our method uses a variation of the great deluge algorithm until a sequence of particular fitness in merit factor is found, indicating that a cluster of high merit factors may be likely. A genetic type algorithm is then used in first iterating through the coordinates to find further generations of sequences of like or better fitness. The next step is to strip these sequences back at either end while fitness is maintained. The genetic method is used to augment these shorter sequences to other new generations, again, as long as fitness is maintained. This fitness level is carefully tuned according to the rate at which high merit sequences are generated. We have found the efficiency of the stochastic algorithm is considerably enhanced by searching through carefully chosen clusters more thoroughly.

### 2.5 Application of inverse collectors problem for statistical analysis

At lengths into the 90s for general sequences and the 180s for skew-symmetric, large numbers sequences with high merit factor are generated, with considerable repetition. We apply results from an analysis of the inverse collectors problem to estimate the number of sequences with  $F > 7$  or 8, say, and to then determine the likelihood that the optimal sequence at a particular length has been found. In this way we establish high confidence that we have the optimal sequences for lengths up to about 90 for general and 180 for skew-symmetric sequences.

## 3 Results

### 3.1 Merit factor records

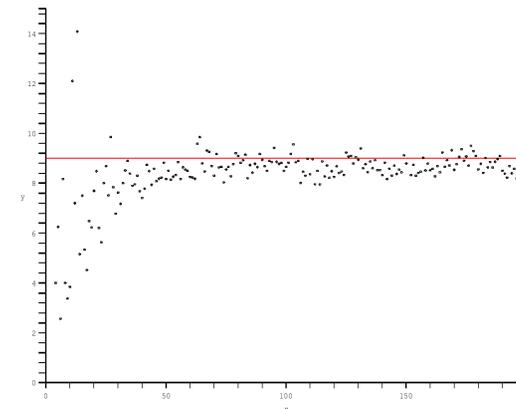


Figure 1: Record merit factors vs. length

Figure 1 displays the largest merit factor found as a function of length. A line at  $F = 9$  is drawn for comparison. The last record found using the general search is at 112. Beyond this all records found are derived from an initial skew-symmetric search. For general sequences, the tendency for the record to approach or surpass 9 is evident at perhaps 100. For skew-symmetric this is observed at about 175.

### 3.2 Asymptotics

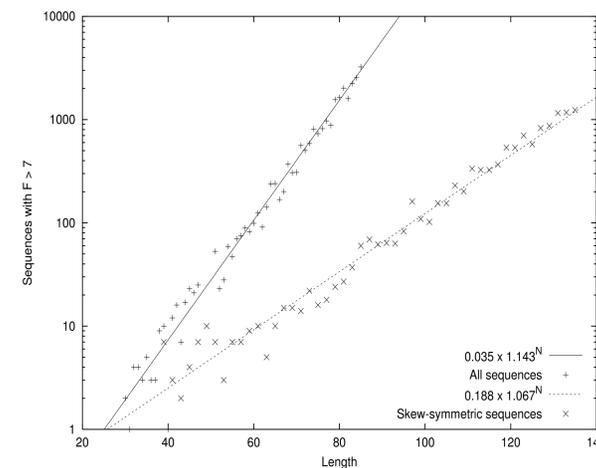


Figure 2: Semilog graphs of numbers of sequences with  $F > 7$

Figure 2 shows semilog plots showing growth is the numbers of sequences with  $F > 7$  for both general and skew-symmetric sequences.

The linear appearance of these plots points to an exponential growth rates. The numbers collected in the upper ranges (800-1000) are sufficient to show an established trend.

In Figure 3 appear estimates for numbers of skew-symmetric sequences with  $F > 8$  and  $F > 8.5$  for lengths between 159 and 183. These suggest exponential growth rate, at least for  $F > 8$  and probably for  $F > 8.5$  as well. We have found an additional 23 sequences with  $F > 9$  to go with 15 previously known, the longest at 102 for the general search and 189 for skew-symmetric.

$N$	$F > 8$	$F > 8.5$
159	44	5
161	57	4
163	52	5
165	57	3
167	73	5
169	73	11
171	58	7
173	90	13
175	97	8
177	99	10
179	153	9
181	114	11
183	125	9

Figure 3: Estimated numbers of skew symmetric sequences

## 4 Conclusions

We have combined stochastic optimization procedures with non-stochastic genetic type methods to first locate clusters of binary sequences with good fitness requirement and then to search these effectively to find their likely optimal examples. This has enabled us to collect enough data to project approximate growth rates for sequences with  $F > 7$  and even  $F > 8$ .

Legendre sequences rotated by 1/4 of their length are known to have an asymptotic limit of  $F = 6$ . A variation proposed by Borwein, Choi, and Jedwab involving extending these at one end by adding a length from the other seems to consistently increase this to about 6.32. Our results suggest an asymptotic greater than 9 developing, and we suspect that in lengths of perhaps 150 for general and 250 for skew-symmetric, that merit factors  $> 10$  will regularly start to appear. However, these are well beyond lengths for which searches can be considered in any way comprehensive using current methods and equipment.